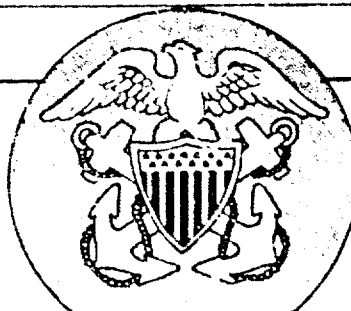


AD 619 692



DDC

AUG 30 1965

DDC-IRA E

# NAVAL MEDICAL FIELD RESEARCH LABORATORY

COPY 2 OF 3  
200  
0.50

Vol. XV, No. 21

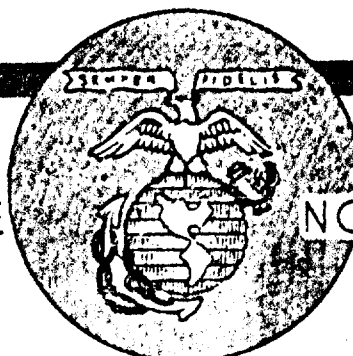
288 July 1965

## A CRITICAL ANALYSIS OF THREE PHYSICAL FITNESS TESTS

by

Philip J. Rasch, Ph. D., and Captain Mark Brown, USMC

Bureau of Medicine and Surgery, Navy Department  
MF022.01-04-8002.2



CAMP LEJEUNE

NORTH CAROLINA

Distribution of this document is unlimited.

ARCHIVE COPY

**A CRITICAL ANALYSIS OF THREE PHYSICAL FITNESS TESTS**

by

**Philip J. Rasch, Ph. D., and Captain Mark Brown, USMC**

---

**U. S. NAVAL MEDICAL FIELD RESEARCH LABORATORY  
CAMP LEJEUNE, NORTH CAROLINA**

**Bureau of Medicine and Surgery, Navy Department  
MF022.01.04-8002.2**

---

**Distribution of this document is unlimited.**

**Submitted by:**

**P. J. RASCH, Ph. D.  
Chief  
Physiology Division**

**Approved by:**

**J. H. BOYERS  
CAPT MC USN  
Commanding Officer**

## **SUMMARY PAGE**

### **THE PROBLEM**

To evaluate the Initial Strength Test, the Physical Readiness Test, and selected Fleishman tests as meaningful measures of the type of fitness required by Marine Corps combat troops.

### **FINDINGS**

---

(1) None of these tests appear entirely satisfactory for this purpose. (2) Types and levels of fitness required by combat troops have never been defined. (3) Satisfactory tests cannot be developed until performance criteria have been established.

### **CONCLUSIONS**

Further work must be undertaken before valid measures of combat fitness can be developed.

### **ADMINISTRATIVE INFORMATION**

Bureau of Medicine and Surgery, Navy Department Research Project MF022.01.04-8002, Report No. 2. Interim report. Approved for publication 23 June 1965.

The authors express appreciation to Major Robert J. Norton, USMC, for his assistance in many phases of this study. The assistance of Lt. I. Dodd Wilson, MC, USNR, Jefferson W. Hamby, HMC, USN, and Raburn Jackson is also acknowledged.

Published by the U. S. Naval Medical Field Research Laboratory, Camp Lejeune, North Carolina.

### **FOR OFFICIAL USE ONLY**

This restriction will be removed and the report may be released on  
15 September 1965

## **ABSTRACT**

The purpose of this investigation was to evaluate the Initial Strength Test, the Physical Readiness Test, and selected Fleishman tests as measures of the type of physical fitness required by Marine Corps combat troops. It was concluded that none of them were satisfactory for this purpose. There are at present no criteria establishing the type and level of fitness required by such troops. Satisfactory tests cannot be developed until such criteria have been established.

---

## A Critical Analysis of Three Physical Fitness Tests

### INTRODUCTION

The investigation of physical fitness in the armed services involves three parameters: (1) the determination of the kind of fitness required of combat troops, (2) the development of satisfactory tests for evaluating this type of fitness and changes therein, and (3) the devising of the most effective methods of improving military physical fitness. At the present time, at least three separate tests for the determination of fitness are in use in the Marine Corps. The physical condition of male Marines under 40 years of age is judged by means of the Physical Readiness Test; the Recruit Training Regiments employ the Initial and Final Strength Tests; the Naval Medical Field Research Laboratory is using a battery based on the factor analysis studies made by Edwin A. Fleishman under a grant from the Office of Naval Research. In the case of the first two norms, standard deviations, percentiles, reliability coefficients, and similar data are not available. Under such conditions, the level of fitness credited to a given individual or the results attributed to a given training program may actually be a reflection of the test used as a criterion. Conceivably a different opinion might have been rendered if the investigators had chosen to use one of the other tests. An essential preliminary to sound work in the field of military physical fitness is a careful examination of the tools by which it is measured and the relationships existing between them. It was the purpose of this study to make a critical examination of the three tests mentioned above.

#### I. Physical Readiness Test

This is by definition a test to determine whether an individual meets certain minimum acceptable standards. It is performed in utilities with boots and helmet, light marching pack, and organic weapon and belt.<sup>1</sup> Since the men with superior fitness have no incentive to perform the test other than in the easiest manner possible, it does not generate data by which the individual can be compared with himself or by which one group can be compared with another.

Event #1 in the test, climbing uphill, consists of stepping on and off a platform 18 inches high for 60 up and down steps in 3 minutes. It is stated that "this event simulates marching uphill at a rapid and steady rate." Whether this is actually the case seems open to some question.

The event is clearly derived from the Harvard Step Test, which purports to measure an individual's "general fitness for hard work."<sup>2</sup> It has been widely used, and equally widely praised and criticized. U. S. Army investigators reported that the test "is a useful one and serves to give an approximate overall evaluation of the fitness of a group of men."<sup>3</sup> U. S. Navy evaluators stated the test "affords a convenient and reliable method for estimating the progress of physical conditioning and of the degree of improvement in such a program."<sup>4</sup>

On the other hand, Cureton and his co-workers have complained that the criterion used for its validation "is expressed in an illegitimate statistical form," includes dependent variables, and is probably invalid.<sup>5</sup> Montoye found there is some slight relationship between Step Test scores and work capacity but it is of little practical importance.<sup>6</sup> Henry and Berg concluded that "physical fitness of the type produced by a typical athletic training regimen can be measured . . . only to a limited extent by performances such as . . . stool stepping to exhaustion."<sup>7</sup>

A test of physical condition is of value only if it has been demonstrated that it correlates highly with physiologic performance in the event which it is desired to test. Attempts to correlate the Harvard Step Test with measures of performance or indices of physique have in general given figures too low to be of predictive value. Some typical examples drawn from the literature are given in Table 1. All of this raises considerable question as to the value of this test in predicting military fitness.

**Table 1**  
**Correlation of Harvard Step Test Scores with**  
**Various Other Tests**

| <b>Event</b>                             | <b>Correlation (<u>r</u>)<br/>with HST</b> | <b>Source<br/>of Data</b> |
|--|--|---------------------------|
| <b>Army Air Forces' Test</b>             | 0.24                                       | 3                         |
| <b>Army Ground Forces' Test</b>          | 0.26                                       | 3                         |
| <b>Mile Run</b>                          | 0.310                                      | 8                         |
| <b>Cross-country Run (1-3/4 mi.)</b>     | 0.38                                       | 6                         |
| <b>Three-mile Run with Marching Pack</b> | -0.21                                      | 9                         |
| <b>Ruffier Index</b>                     | -0.39                                      | 10                        |
| <b>Pignet Index</b>                      | 0.14                                       | 10                        |
| <b>Reciprocal Buffon Index</b>           | 0.13                                       | 10                        |
| <b>Bruce Physical Fitness Index</b>      | 0.236                                      | 11                        |

While many additional correlations of this test with various other criteria have been reported in the literature, the writers have been unable to find any occasion on which it has been used to evaluate ability to move uphill. However, the fact that it has a low correlation with cross-country running, which ordinarily includes a good deal of uphill running, may be significant. Rovelli and Aghano<sup>12</sup> specifically deny that it can be used as a test of ability to develop maximal energy expenditure, as in running uphill.

The Harvard Step Test has no significant correlation with height or combined height-weight factors, although extremely heavy men may make relatively low scores.<sup>13</sup> However, Reibourn<sup>14</sup> has commented that the fine scores made by Gurkha troops in the Harvard Pack Test, which is also one of rapid step climbing, appeared to be related to their exceptional calf development. It remains to be determined whether the step test as used in the Physical Readiness Test is also related to calf development.

The muscle action involved in this test suggests that it might prove a valid measure of the ability to move through deep mud, such as is said to characterize the rice paddies of Viet Nam.

Event #2 in this test is a 20-foot rope climb. This item certainly has "face validity," but there seems to be little information on precisely what it measures. Fleishman<sup>15</sup> submitted a somewhat similar test to factor analysis and reported a loading of 0.67 with dynamic strength and 0.41 with explosive strength. In all probability, a quite similar loading would be found for the Marine Corps version of the rope climb.

Event #3, evacuation, requires a man to run 50 yards in a sig-sag fashion, lift a "casualty," and carry him back to the starting point. The fact that the sig-sag is not according to a specified plan introduces an uncontrolled variable. The fact that the weight of the "casualty" may vary on each test is a second uncontrolled variable.

Event #4, advance by fire and maneuver, requires that the Marine "creeps or crawls" for 25 yards, thus immediately introducing an uncontrolled variable. He is then to run in a sig-sag fashion, thereby introducing a second uncontrolled variable.

Since both Events #3 and #4 include two uncontrolled variables, it is evident that there may be considerable difference in the way in which they are run by different individuals or even by the same individual on different occasions. It would be expected that test-retest reliability would be undesirably low.

Event #5, forced march, requires a 3-mile run-walk carrying a light marching pack. Like Event #2, this item has face validity.

It will be noted that Events #3, #4, and #5 all require the subject to run with a load of some kind. This suggests the possibility that if each of these events were standardized and run for time, a high intercorrelation would be found between them. In that case the inclusion of these three events instead of only the best of them adds little or nothing to the test battery.

Another difficulty is that events of this type may be quite seriously affected by the terrain. Scores on a dry, level, firm surface should be much lower than those in muddy, sandy, or hilly country. From the standpoint of practical administration, further difficulties are often encountered. The 3-mile forced march is sometimes made in platoon formation, rather than as an individual effort. The individual events are often given in whatever order is most convenient, with the exception that the 3-mile forced march is normally the final event in the series. The difficulty is that each event affects the scores in those which follow it. Unless the order of administration is the same each time, the scores cannot be compared.

The minimum requirements for any satisfactory test are that it be valid, reliable, standardized, and normed for the population being tested.\* Since these requirements are not met by the Physical Readiness Test, it is impossible to use scores based upon it for any sort of statistical analysis.

## II. Initial and Final Strength Tests

The physical fitness of recruits at the Recruit Training Regiment, Parris Island, is evaluated at the start of their training by means

---

\* Reliability refers to the ability of a set of measurements to give consistent results. The reliability of a certain instrument applies to a certain population under certain conditions. It is usually reported in terms of a reliability coefficient which expresses the relationship between two measurements. A test is valid if it measures what it purports to measure. This is usually stated as a validity coefficient, which expresses the relationship between the predictor and the criterion. For technical reasons, reliability and validity require opposite approaches in test construction. The result is that the two can never both be maximal in a single test. In actual practice, a tester may seek to combine several reliable tests into a valid battery.



of the Initial Strength Test and at the end by means of the Final Strength Test.<sup>16</sup> In both tests they wear gym suits and sneakers. The items in this battery are varied according to the weather. If the weather is suitable, the recruits are required to perform the following:

1. Pull-ups, with palms out
2. Push-ups
3. Sit-ups
4. Bend and thrust
5. 300-yard shuttle run (60 yards x 5)

---

In event of inclement weather, the 300-yard shuttle run is replaced by side straddle hops, so that actually two forms of this test exist. Apparently this battery dates back to World War II and is identical with that used by the Army.<sup>17</sup> No intercorrelations of the test items, means, standard deviations or norms for the scores made by contemporary Parris Island Marine Corps recruits have been found in the available literature. Therefore it was first necessary to determine these statistics. Data collected on a representative number of recruits were analyzed\* and the findings are shown in Table 2. The combined group consists of 248 men who routinely performed the 300-yard shuttle run as part of their test battery and 256 men (Group B) who were tested during foul weather and would normally have performed only the side-straddle hops. Since it was impossible to determine the correlation between these two items from what are normally dichotomous groups, Group B was required to perform both of these tests within a day or two of each other. For comparative purposes, similar data recorded by Bates<sup>18</sup> at the San Diego Marine Corps Recruit Depot in 1959 are also displayed. No figures are shown under bend and thrust and side straddle hops, since the San Diego testers used the squat-jump in place of these. Simple inspection indicates that there are no great differences between the two groups in the scores for the other events.

Reference to the scoring table for the Initial Strength Tests, however, raises certain questions. Presumably the mean figures represent a point of equal difficulty in each case, and it would be anticipated that the same number of points would be awarded in each case

---

\* The authors are indebted to Lt. Col. C. R. Livingston, USMC, Data Processing Officer, Data Processing Installation No. 2, Marine Corps Base, Camp Lejeune, and to Capt. E. J. Doran, USMC, Assistant Data Processing Officer, for designing programs and processing most of the statistical work in this report.

**Table 2**  
**Means and Standard Deviations of Initial Strength Test Items**

| Group     | N*    | Pull-ups |     | Sit-ups |      | Bend-Thrust |     | Push-ups |      | 300-yd Shuttle Run (sec) |     | Side Straddle Hops |      |
|-----------|-------|----------|-----|---------|------|-------------|-----|----------|------|--------------------------|-----|--------------------|------|
|           |       | M†       | SD‡ | M       | SD   | M           | SD  | M        | SD   | M                        | SD  | M                  | SD   |
| Combined  | 504   | 5.7      | 2.9 | 46.0    | 11.0 | 24.0        | 5.3 | 27.4     | 10.6 | 57.4                     | 4.9 |                    |      |
| Group B   | 256   |          |     |         |      |             |     |          |      |                          |     |                    |      |
| San Diego | 9,421 | 5.1      | 3.1 | 41.4    | 10.4 |             |     | 24.0     | 8.3  | 53.9                     | 3.7 | 143.3              | 17.0 |

\* N = Number.

† M = Mean.

‡ SD = Standard deviation.

for attaining it. As is shown in Table 3, this is not the case. Actually it appears that the bend and thrust is under-evaluated and the side straddle hops over-evaluated.

From the standpoint of test administration, scoring tables of this type are undesirable, since they yield numerical values which have no particular significance to either the testee or the tester. Guilford<sup>19</sup> recommended centile rank positions\* as the most meaningful to the non-statistician, and there would seem to be little reason to disagree with his suggestion. Adoption of such a method of scoring would have the distinct advantage that it would be relatively simple to place a profile chart on the back of the recruit's score card. This would be of great assistance in evaluating both the recruit's improvement and the effectiveness of the training program itself.

The scores for each test were intercorrelated by means of the Pearson  $r$ . The results are shown in Table 4. Since four of the tests were common to both groups, the scores were combined and the overall  $r$  computed. Examination of these figures reveals that they are

Table 3

Means and Equivalent Points for Initial Strength Test Items

| Test               | Mean | Points |
|--------------------|------|--------|
| Pull-ups           | 5    | 45     |
| Sit-ups            | 46   | 49     |
| Bend and Thrust    | 24   | 32     |
| Push-ups           | 27   | 50     |
| 300-yd Shuttle Run | 57   | 53     |
| Side Straddle Hops | 143  | 68     |

---

\* A centile is a point on a scoring table below which is any given proportion of scores. That is, 79% of the population will attain a score less than that represented by the 80 centile.

**Table 4**  
**Intercorrelations Between Initial Strength Test Items**

| Test Item                 | N   | Sit-ups | Bend-Thrust | Push-ups | 300-yd Shuttle Run | Side Straddle Hops |
|---------------------------|-----|---------|-------------|----------|--------------------|--------------------|
| <b>Pull-ups</b>           |     |         |             |          |                    |                    |
| Group A                   | 248 | 0.16    | 0.23        | 0.53     | -0.05              |                    |
| Group B                   | 256 | 0.25    | 0.17        | 0.48     |                    | 0.19               |
| Combined                  | 504 | 0.21    | 0.21        | 0.52     |                    |                    |
| <b>Sit-ups</b>            |     |         |             |          |                    |                    |
| Group A                   | 248 |         | 0.22        | 0.27     | -0.11              |                    |
| Group B                   | 256 |         | 0.12        | 0.24     |                    | 0.19               |
| Combined                  | 504 |         | 0.18        | 0.27     |                    |                    |
| <b>Bend &amp; Thrust</b>  |     |         |             |          |                    |                    |
| Group A                   | 285 |         |             | 0.27     | -0.10              |                    |
| Group B                   | 256 |         |             | 0.37     |                    | 0.25               |
| Combined                  | 541 |         |             | 0.30     |                    |                    |
| <b>Push-ups</b>           |     |         |             |          |                    |                    |
| Group A                   | 285 |         |             |          | -0.01              |                    |
| Group B                   | 256 |         |             |          |                    | 0.29               |
| <b>300-yd Shuttle Run</b> |     |         |             |          |                    |                    |
|                           | 256 |         |             |          |                    | -0.21              |

satisfactorily orthogonal\* with the possible exception of pull-ups versus push-ups, where a combined correlation of  $r = 0.52$  was obtained. Working with Great Lakes naval recruits, Fleishman obtained an almost identical correlation ( $r = 0.58$ ), and commented that 'push-ups added to pull-ups contributes little new information regarding a subject's dynamic strength.'<sup>20</sup> The same reasoning seems applicable here.

A point of special interest is found in the fact that the correlation between the 300-yard shuttle run and the side straddle hops is so low. Whatever the latter measures, it is not the same factor measured by the former. So far as this item is concerned, the recruits received during foul weather are given a different test than are those received during good weather, with the further advantage that points are easier to earn in the side straddle hop than in the 300-yard shuttle run.

### III. Fleishman Tests

In 1958 the Office of Naval Research initiated a project entitled "The Development of Criteria of Physical Proficiency." This was assigned to Yale University and was directed by Edwin A. Fleishman.<sup>15</sup> A factor analysis of the findings of previous research in the field identified 14 factors of physical proficiency. The final outcome was a proposed Fitness Test battery designed to measure 11 factors: explosive strength, static strength, dynamic strength, trunk strength, extent flexibility, dynamic flexibility, gross body equilibrium, balance, speed of limb movement, gross body coordination, and stamina. The present investigators considered that five of these items were of special interest for the testing of combat troops: explosive strength (ability to exert maximum energy in one explosive act), static strength (exertion of a maximum force for a brief period of time), dynamic strength (strength of limbs in moving or supporting the weight of the body repeatedly over a period of time), trunk strength, and stamina (cardiovascular endurance during prolonged exertion of the body). Fleishman recommends the following as the respective tests of choice for each of these factors: shuttle run (5 x 20 yards), hand grip with dynamometer, pull-ups, leg lifts (maximal number in 30 seconds), and 600-yard run-walk. For the purpose of this paper these will be referred to as the Fleishman Tests. Norms and centiles are available for each test.<sup>15</sup>

---

\* The different items measure different qualities. High correlations between separate events indicate that the tests simply measure the same thing in different ways.

Experience with this battery revealed certain problems in its administration. On the basis of its use at both Parris Island and Camp Lejeune, the shuttle run does not appear satisfactory. The ground at these bases is soft, sandy, and grassy. A number of subjects slide or slip and fall when reversing the direction of the run. The first few testees pound a hole into the ground at the point where the change of direction takes place. The following runners simply plant their foot in this hole and pivot, thereby gaining a mechanical advantage which obscures differences in actual speed and agility.

---

Reference to Fleishman's book shows that his second choice for the measurement of explosive strength is the 50-yard dash. This correlates  $r = 0.80$  with performance in the shuttle run and has a loading of  $r = 0.75$  with the factor identified as explosive strength. This does not appear to be essentially different from the loading of  $r = 0.77$  found for the shuttle run. The reliability of the two tests is practically identical:  $r = 0.86$  for the 50-yard dash and  $0.85$  for the shuttle run. In view of the close relationship between the two events and the demonstrated unsatisfactory nature of the shuttle run under our conditions, experiments will be made in substituting the 50-yard dash for the shuttle run in future studies.<sup>21</sup>

Fleishman concluded that the pull-up was the best measure of dynamic strength and that use of the "under-hand grip" (palms facing the subject) was preferable to the "over-hand grip" (palms facing away from the subject), since more pull-ups can be done this way and a better distribution of scores is obtained. He gives no comparative figures on the two methods, but his opinion agrees with an earlier Army Air Force statement to the effect that chinning performances are superior when the under-hand grip is used,<sup>22</sup> and has received support from an electromyographic study of the muscles involved.<sup>23</sup> This raises a problem in that the Initial Strength Test specifies that pull-ups are to be done with the palms out.

From the information available in the literature, there appears to be relatively little difference in the scores for the two styles. DeWitt,<sup>24</sup> using college men as subjects, obtained a mean of 9.71 for the under-hand style and one of 7.63 for the over-hand grip, as a difference of 2.08 in favor of the former. Experienced testers agree that unless otherwise directed most of the subjects will choose the former. It has been suggested that this is because the over-hand grip seems more fatiguing, although it is not actually more costly in energy.<sup>25</sup> However, when the test is administered to Marine Corps personnel in this fashion, the officers and non-commissioned officers in charge object to the use of the under-hand grip. It is their contention that in scaling

a wall, swinging into a window, or ascending a roof, the trooper must use the over-hand grip, and that both as a matter of training and in fairness to the individual he should be tested as he will perform.

It is, of course, not necessary that the test duplicate the criterion; only that a satisfactorily high validity correlation exists between the two. The correlation between these two forms of doing pull-ups and between test-retest scores when Army Air Force cadets are used as subjects<sup>26</sup> has been shown to be on the order of  $\bar{r} = 0.735$  to  $0.795$ . McGraw,<sup>27</sup> also studying college men, found means of 9.94 and 8.02 respectively on one occasion and of 10.84 and 9.82 respectively on a second occasion, from which he concluded that while the under-hand grip gave the higher scores, the day-to-day variations ( $d = 0.90$  and  $1.80$  respectively) were apt to be as large as the difference between grips ( $d = 1.92$  and  $1.02$  respectively). McGraw reports a test-retest coefficient of  $\bar{r} = 0.73$  for the under-hand grip and  $0.88$  for the over-hand grip, remarking that the former figure "is well below the value usually accepted for retest reliability." The coefficient of  $\bar{r} = 0.73$  for the under-hand grip is somewhat surprising, as Fleishman found that pull-ups had test-retest reliability of  $\bar{r} = 0.93$  when used with recruits at Great Lakes Naval Training Center.

It is not clear whether the Army Air Force cadets, college men and Navy recruits are from the same population as are Marine Corps combat troops. To clarify this point, it was necessary to determine the correlation of the two styles and the test-retest reliability of this item when used with typical Marine Corps infantrymen.

Forty-eight men from "G" Company, 2d Battalion, 2d Marine Division, stationed at Camp Lejeune, served as volunteer subjects.\* However, only 31 completed all tests and are reported on here. Pull-ups were included in their routine daily physical training program, so that the problem of muscle soreness or of fatigue from unaccustomed exercise was not a factor in the findings.

On 26 January 1965 half of the subjects performed chins using the under-hand grip. The other half used the over-hand grip. The following day the tests were repeated, with the men reversing their grip. On the third and fourth days respectively the testing program of the first and second days was replicated. A half point was counted if a man could get his upper arms parallel to the ground although he could not get his chin over the bar.

---

\* The authors are indebted to 1st Lt. F. Leroy Scovill, III, USMC, of "G" Company, for his cooperation in this phase of the study.

A number of drill instructors had commented that pull-up scores appeared to be inversely related to the man's body weight. It is almost self-evident that this will be true when the body weight includes a high percentage of fat. The relationship between these two variables when all subjects are in a state of vigorous physical training and their weight is presumably predominantly lean body mass is less apparent. To answer this question, the body weight of the subjects was also recorded on the occasion of the first test.

During the tests it was determined that the men included rope climbing in their physical fitness training program. Advantage was taken of this to determine the correlation (Pearson  $r$ ) between pull-ups, body weight and rope climbing, with and without packs, since there is a "face validity" between pull-ups and rope climbing. On 4 February 1965 the subjects performed the rope climb in accordance with the instructions laid down in the Marine Corps Physical Fitness Test.<sup>28</sup> The rope was 20 feet long, 1-1/2 inches in diameter and knotted approximately every 2 feet. The man started in the standing position and grasped the rope as high as he could reach. The stop watch was started when the command "Go" was given and stopped when the man touched the beam from which the rope was suspended. Half of the men climbed with utilities, boots, helmet, light marching pack, and organic weapon and belt; the other half wore utilities and belts only. The following day this was reversed, so that those who had climbed with helmets, pack and weapon climbed with utilities and belts only, and the others climbed with the required gear. The position of the first two or three knots in relation to the height of the man makes for differences in starting styles. A tall man may be able to reach above a knot and secure a comfortable grasp prior to starting; a short man may have to await the starting signal and then make a little jump to secure his handhold. For this reason, times were taken only to the nearest 0.5 second.

During this whole series of tests the weather was in the low thirties. The men's hands were chilled, the pull-up bar was cold, and the ropes were not only cold but stiff. As a result, scores are probably lower and times higher than would have been the case in more moderate weather.

The mean data for the two methods of performing the pull-ups are shown in Table 5. For comparative purposes the mean scores of the first pull-up attempts of all groups reported by the investigators cited above are shown in Table 6.

Inspection of these data would suggest that there is comparatively little difference between aviation cadets, college men, and Marine



Table 5

## Intercorrelations of Pull-ups

| Trial       | N  | Over-hand Grip |     | Under-hand Grip |     | Correlation |
|-------------|----|----------------|-----|-----------------|-----|-------------|
|             |    | Mean           | SD  | Mean            | SD  |             |
| First       | 49 | 7.8            | 3.0 | 9.9             | 3.0 | 0.96        |
| Second      | 49 | 7.6            | 2.7 | 10.2            | 3.0 | 0.89        |
| Correlation |    | 0.95           |     | 0.93            |     |             |

Table 6

## Comparison of Pull-up Scores of Various Groups

| Subjects                      | N    | Over-hand Grip |      | Under-hand Grip |      |
|-------------------------------|------|----------------|------|-----------------|------|
|                               |      | Mean           | SD   | Mean            | SD   |
| Aviation Cadets <sup>26</sup> | 4057 | 8.52           | 3.09 | 9.45            | 3.29 |
| Aviation Cadets <sup>26</sup> | 3445 | 8.28           | 2.96 | 9.17            | 3.17 |
| College Men <sup>24</sup>     | 144  | 7.63           |      | 9.71            |      |
| College Men <sup>27</sup>     | 51   | 8.02           |      | 9.94            |      |
| Navy Recruits <sup>29</sup>   | 201  |                |      | 5.96            | 3.61 |
| Marine Troops                 | 49   | 7.80           | 3.00 | 9.90            | 3.00 |

Corps troopers, but the stated figures very likely underestimate the comparative abilities of the Marines. These men were tested while wearing fatigue clothes and field boots. It is assumed that all other groups were tested in gym costume and tennis shoes. Quite likely they also had the additional advantage of more favorable environmental conditions. Acceptance of the mean figure of 9.9 would place the Marines at the 70th percentile on Fleishman's national norms, but under similar test conditions that would probably rate five or ten points higher than this. In any event, they seem distinctly superior to the Navy

recruits, indicating that norms based on naval personnel are not necessarily valid for Marine Corps troops.

As has been true in all previous studies, our subjects made higher scores with the under-hand grip than with the over-hand grip. What is of primary interest is that the correlations of first under-hand versus first over-hand, first under-hand versus second over-hand, second under-hand versus first over-hand, and second under-hand versus second over-hand are all  $r = 0.89$  or better (Table 5). This indicates that men who do well on one style will also do well on the other. Since the type of grip will have relatively little effect on the relative placement of the individual being tested, it is desirable to use the under-hand grip in fitness testing in order to secure a greater distribution of scores. As a minor benefit, this would make it possible to compare the scores of Marines with pre-determined national norms.<sup>15</sup>

The correlation for first under-hand versus second under-hand is identical with that reported by Fleishman,  $r = 0.93$ , and practically identical with that for first over-hand versus second over-hand,  $r = 0.95$ . This indicates that for subjects accustomed to practicing pull-ups in their regular physical training this test has a high reliability.

The mean body weight of our subjects was 165.6 pounds (S. D. = 18.2 pounds). When the under-hand grip is used, the correlation for the first pull-up scores versus body weight is  $r = -0.09$ ; with the over-hand grip, it is  $r = -0.21$  (Table 7). Thus when dealing with well-conditioned troops, the influence of body weight on the scores is negative, but to such a limited degree as to be of little consequence. There is, then, no need to take body weight into consideration when giving pull-up tests to trained Marine Corps troops. The situation with untrained recruits may be quite different, and requires further study. The picture is almost identical insofar as the effect of body weight on rope climbing is concerned.

The first over-hand grip scores and rope climb with pack correlate  $r = -0.49$ . With the under-hand grip the correlation is  $r = -0.51$ , which is again essentially identical (Table 8). (The negative

Table 7

Intercorrelation of Body Weight, Pull-ups, and Rope Climb (N = 32)

| Event       | Over-hand<br>Pull-ups | Under-hand<br>Pull-ups | Rope Climb<br>without Pack | Rope Climb<br>with Pack |
|-------------|-----------------------|------------------------|----------------------------|-------------------------|
| Body Weight | -0.21                 | -0.09                  | -0.17                      | -0.01                   |

Table 8

## Intercorrelations of First Pull-ups and Rope Climb

| Style of Pull-up | N  | Rope Climb<br>without Pack<br>(sec) |     | Correla-<br>tion | Rope Climb<br>with Pack<br>(sec) |     | Correla-<br>tion |
|------------------|----|-------------------------------------|-----|------------------|----------------------------------|-----|------------------|
|                  |    | Mean                                | SD  |                  | Mean                             | SD  |                  |
| Over-hand Grip   | 31 | 11.8                                | 3.2 | -0.58            | 17.5                             | 5.4 | -0.49            |
| Under-hand Grip  | 31 |                                     |     | -0.59            |                                  |     | -0.51            |

correlation here may be confusing at first sight, but reflects the fact that in rope climbing a decreased time constitutes a superior performance.) This shows a moderate but substantial relationship between performance in the two events. The coefficient of determination\* must then be on the order of 25% which indicates that the two events are not entirely orthogonal. With a correlation as great as  $r = 0.50$ , little additional information would be gained by including both items in a single test battery. However, the proportion that is independent is so large that performance in one will not serve to predict satisfactorily performance in the other. This confirms Fleishman's statement that the rope climb has a large factor loading with dynamic strength.

While Fleishman recommends use of the 600-yard run-walk as a measure of stamina, this seems to have been an afterthought. This event is not included in his tables of intercorrelations between tests. Hence the first step was to determine whether it was in fact orthogonal to the other events in the Fleishman battery. Data were therefore collected on the test scores of 275 recruits at Parris Island early in 1965. The intercorrelations are displayed in Table 9, from which it is clear that this test is orthogonal to the other items.

Even with this established, another problem was evident. The Initial Strength Test employs the 300-yard shuttle run as a measure of cardiorespiratory endurance, the Fleishman tests utilize the 600-yard run-walk, and the Physical Readiness Test incorporates the 3-mile

---

\* The coefficient of determination is defined as  $r^2$ . It represents the percentage of individual differences in one variable which is associated with or determined by the individual differences in another variable.

Table 9  
Intercorrelations Between Selected Test Items  
in Fleishman Tests (N = 275)<sup>30</sup>

| Test Item   | Hand Grip | Shuttle Run | Leg Lift | 600-yd<br>Run-Walk |
|-------------|-----------|-------------|----------|--------------------|
| Pull-ups    | 0.07      | -0.07       | 0.40     | -0.33              |
| Hand Grip   |           | -0.08       | 0.11     | -0.18              |
| Shuttle Run |           |             | -0.17    | 0.15               |
| Leg Lift    |           |             |          | -0.33              |

forced march. It is not self-evident that these measure the same factor in fact, Cureton<sup>31</sup> apparently considers that the 300-yard shuttle run is a measure of explosive strength rather than of cardiorespirator endurance. To clarify this point, 61 Camp Lejeune Marines served as subjects. The prescribed gear was worn during the 3-mile force march. The other two tests were performed in gym suits and sneakers. The men ran each of the three events on a Latin square design and the times were intercorrelated. The means and standard deviations are shown in Table 10. Intercorrelations between the three scores are displayed in Table 11.

Table 10  
Means and Standard Deviations of Running Events  
(N = 61)

| Event              | Mean     | SD      |
|--------------------|----------|---------|
| 300-yd Shuttle Run | 52.7 sec | 2.3 sec |
| 600-yd Run-Walk    | 89.8 sec | 4.4 sec |
| 3-mi Forced March  | 32.0 min | 2.2 min |

Table 11  
Intercorrelations of Running Events (N = 61)

| Test Item       | 600-yd Run-Walk | 3-mi Forced March |
|-----------------|-----------------|-------------------|
| Shuttle Run     | 0.67            | 0.17              |
| 600-yd Run-Walk |                 | 0.34              |

It is clear that the 300-yard shuttle run and the 600-yard run-walk are substantially related, and that neither of them are of any value in predicting performance in the 3-mile forced march. It was considered possible that the load represented by the gear required in the Physical Readiness Test might have some influence on the correlations between the 600-yard run-walk and the 3-mile forced march. To test this hypothesis, 10 of the original 61 subjects ran the same 600-yard course again, this time dressed as prescribed by the Physical Readiness Test. Their times were correlated with those previously determined. The means and standard deviations are displayed in Table 12 for the 3-mile forced march and the correlation ( $r = 0.35$ ) is shown in Table 13. Since the addition of the load made no essential difference in the correlation, it strongly suggests that the essential point involved is the distance. It would appear that both the 300-yard shuttle run and the 600-yard run-walk are largely measures of explosive strength. If the 3-mile forced march is a valid criterion of the type of cardiorespiratory endurance desirable in the Marine Corps combat infantryman, neither the Initial Strength Test nor the Fleishman Test battery includes an item which measures this parameter.

**Table 12**

**Means and Standard Deviations of 600-yard Run-Walk (sec)**

| Conditions     | N  | Mean  | SD  |
|----------------|----|-------|-----|
| In gym costume | 61 | 89.8  | 4.4 |
| In light pack  | 10 | 111.2 | 7.8 |

**Table 13**

**Correlation of 600-yard Run-Walk in Light Combat Pack  
and 3-mile Forced March  
(N = 10)**

| Event               | Mean      | $r$  |
|---------------------|-----------|------|
| 600-yard Run-Walk   | 111.2 sec | 0.35 |
| 3-mile Forced March | 32.7 min  |      |
| Correlation         |           |      |

Work lasting up to approximately 1 minute is said to depend largely on anaerobic work capacity (ability to liberate energy in the absence of oxidation) while longer periods are controlled by aerobic work capacity. Perhaps the ability to perform work under aerobic conditions cannot be predicted by tests completed, or largely completed, under anaerobic conditions. If so, it may be expected that a run of at least 1/2 mile will be required to predict the time of the 3-mile forced march to any usable degree.

On the basis of Fleishman's work, it would appear that the events comprising the Initial Strength Test may be classified as follows:

|                      |                               |
|----------------------|-------------------------------|
| Pull-ups             | Dynamic strength (arms)       |
| Bend and thrusts     | Dynamic strength (legs)       |
| Push-ups             | Dynamic strength (arms)       |
| Sit-ups              | Trunk strength (weak measure) |
| 300-yard shuttle run | Explosive strength            |
| Side straddle hops   | ?                             |

The primary problem is to determine just what kind and how much fitness a combat Marine needs. The writers have heard one officer argue that he actually needs very little, because most of his time is spent crouching in a shell hole, from which he emerges only to run a few yards to another protected spot. Unquestionably, much of the fatigue of combat is psychological, resulting from fear, hunger, shock, panic, mental fatigue, and loss of sleep.<sup>32</sup> The extent to which these can be offset by physical conditioning is unknown. It is quite possible that the problem is primarily one of getting men to a given area in condition to fight. The British Royal Marines use speed marching, in which a man is required to cover up to 9 miles at the rate of a mile in 10 minutes as one of their criteria.<sup>33</sup> A British Royal Marine captain now at Camp Lejeune has informed one of the writers that on the basis of his experience in three different campaigns he considers this quite a satisfactory measure of cardiorespiratory fitness for combat. In his personal opinion, the 3-mile forced march is not satisfactory for this purpose. It will be noted that the mean time of our 61 subjects in the 3-mile forced march was 32.0 minutes (Table 10). In the opinion of the observers, their condition at the end of this run was such as to render it highly unlikely that they could have sustained this pace for another 6 miles. By British Royal Marine standards, these men would almost certainly require further conditioning. The proper approach to this problem would seem to be the direct one - actually measure a group of men who have demonstrated their fitness by successful participation in arduous combat patrols and similar maneuvers in Viet Nam and determine their performance capabilities.

The British standards are, of course, set for their commando oriented Royal Marines. In the U. S. Marine Corps, which is much larger, we have specialized troops, such as tankers, artillerymen, and supply personnel. It is quite possible that such specialists have no need for the same level of physical fitness that is required by the combat infantrymen. In such case it would be comparatively simple to establish different requirements for different branches of the Corps.

From the standpoint of modern athletic training theory, the ingredient that seems to be commonly missing from the Marines' fitness training is all-out effort. The troops at Camp Lejeune do a great deal of double timing, but it is extremely rare to see them running. However, it is precisely this level of stress which is needed in order to achieve high levels of fitness. It is suggested that attention might well be given to the introduction of interval training into the conditioning program of the Marine Corps.<sup>34</sup> One difficulty with this is that the "all-out" effort of men varies and the troops tend to become so spread out that military control is lost. Some modification would probably be required in order to keep the men under the control of their officers.

### SUMMARY

1. At the present time there is no general agreement as to what kinds of fitness and what levels of fitness are needed by combat troops. Until a decision has been reached on this point, it will be impossible to develop meaningful test batteries.

2. The Physical Fitness Readiness Test makes certain assumptions which are open to question and includes items which may be repetitious. Since it is based on minimum performances and permits differences in administration, it cannot be used to compare groups or to measure changes in condition.

3. The Initial Strength Test has a dry weather and a wet weather battery. The two are not equivalent. The scoring tables are to some extent inequitable. The battery is heavily weighted with dynamic strength tests and lacks static strength and stamina tests. Use of a different technique in the pull-up would improve the distribution of the scores. A different method of scoring would be more informative for both testers and testees.

4. The Fleishman Tests appear to have a sound theoretical basis. Under certain conditions, the 100-yard shuttle run is unsatis-

factory from an administrative viewpoint. The 600-yard run-walk is not a satisfactory measure of cardiorespiratory endurance if the 3-mile forced march is used as the criterion.

#### REFERENCES

1. Marine Corps Order 6100.3C, 29 October 1962.
2. Brouha, L., A. Graybiel and C.W. Heath. The Step Test. Rev. Canad. Biol. 5:86-91, February 1943.
3. Bean, W. B., and others. A Critique of Physical Fitness Tests. Army Medical Research Laboratory Report No. 3, 19 February 1947.
4. Behnke, A. R., and others. The Step-Up Test as a Measure of the Progress of Physical Conditioning. Naval Medical Research Institute Research Project X-134. Report No. 1, 1 March 1943, p. 3.
5. Cureton, T. K., and others. Physical Fitness Appraisal and Guidance. St. Louis: C. V. Mosby Co., 1947, p. 176.
6. Montoye, H. J. The "Harvard Step Test" and Work Capacity. Reproduced from Rev. Canad. Biol. 11:491-499, March 1953.
7. Henry, F. M. and W. E. Berg. Observations on the Harvard Step Test. J. Appl. Physiol. 3:103-111, July 1950.
8. Cureton, T. K., and others. Endurance of Young Men. Washington, D. C.: Society for Research in Child Development, 1945.
9. Rasch, P. J. and I. D. Wilson. The Correlation of Selected Laboratory Tests of Physical Fitness with Military Endurance. Mil. Med. 129:256-258, March 1964.
10. Monod, H. and S. Bouisset. An Evaluation of Three Standard Tests of Physical Fitness. Int. Z. Angew. Physiol. 20:223-232, 21 February 1964.
11. Rasch, P. J., H. S. Cowell, and W. R. Pierson. The Correlation of the Bruce PRI with the Harvard Step Test RI. Rev. Canad. Biol. 18:77-82, April 1959.



12. Rovelli, E. and P. Aghano. Physiological Characteristics of the "Step" Test. Int. Z. Angew. Physiol. 20:190-194, 16 October 1963.
13. Reedy, J. D., G. L. Saiger, and R. H. Holser. Evaluation of the Harvard Step Test with Respect to Factors of Height and Weight. Int. Z. Angew. Physiol. 17:115-119, 1958.
14. Renbourn, E. T. Physiological Problems of the Soldier in Tropical Warfare. Royal Army Medical Corps, 105:172-182, October 1959.
15. ~~Fleishman, E. A. The Structure and Measurement of Physical Fitness. Englewood Cliffs: Prentice-Hall, Inc., 1964.~~
16. Recruit Training Regiment, Parris Island Lesson Plan Course 2000. 30 January 1964.
17. Physical Conditioning. Department of the Army Technical Manual TM21-200, December 1957.
18. Bates, F. M. Physical Fitness of Marine Recruits. Unpublished Masters' Thesis, San Diego State College, August 1960.
19. Guilford, J. P. Fundamental Statistics in Psychology and Education. 2nd Ed. New York: McGraw-Hill Book Co., 1950, p. 289.
20. Fleishman, op.cit., p. 72.
21. Rasch, P. J., and others. Physical Training in Confined Spaces. I. Evaluation of the Universal Gym - A Preliminary Report. NMPRL Report, XV, No. 13, May 1965.
22. Headquarters, Army Air Forces. The Army Air Force Physical Fitness Research Program. Research Quarterly, 15:12-15, March 1944.
23. Randall, Nellie G. An Electromyographic Study of Selected Muscles Involved in Two Methods of Chinning. Unpublished Ph.D. dissertation, State University of Iowa, August 1963.
24. DeWitt, R. T. A Comparative Study of Three Types of Chinning Tests. Research Quarterly, 15:249-251, October 1944.
25. Hamlin, H. E. and F. A. Waterman. The Energy Cost of Chinning the Bar and Push-Up Exercises. Research Quarterly, 7:75-80, October 1936.

26. Phillips, B. E. The JCR Test. Research Quarterly, 18:12-29, March 1947.
27. McGraw, Lynn W. Effect of Variation of Forearm Position in Elbow Flexion. Research Quarterly, 35:504-510, December 1964.
28. Marine Corps Order 6100.3C, 29 October 1962.
29. Fleishman, op. cit., p. 59.
30. Rasch, P. J., and others. Evaluation of an Experimental Combat Conditioning Course and of Physical Fitness Tests. NMFRL Report, XV, No. 6, March 1965.
31. Cureton, T. K., and others. Endurance in Young Men. Washington, D. C.: Society for Research in Child Development, 1945, p. 73.
32. United States Combat Developments Command. A Study to Conserve the Energy of the Combat Infantryman. Fort Benning, September 1963, p. 8.
33. Light, K. E. to R. J. Norton, letter DO/KEL/29/64, 16 September 1964.
34. Morehouse, L. E. and P. J. Rasch. Sports Medicine for Trainers. 2nd Ed. Philadelphia: W. B. Saunders Co., 1963, p. 108-116.